Feature Trajectory Retrieval with Application to Accurate Structure and Motion Recovery

Kai Cordes, Oliver Müller, Bodo Rosenhahn, and Jörn Ostermann

Institut für Informationsverarbeitung (TNT)
Leibniz Universität Hannover
{cordes,omueller,rosenhahn,ostermann}@tnt.uni-hannover.de
http://www.tnt.uni-hannover.de/

Abstract. Common techniques in structure from motion do not explicitly handle foreground occlusions and disocclusions, leading to several trajectories of a single 3D point. Hence, different discontinued trajectories induce a set of (more inaccurate) 3D points instead of a single 3D point, so that it is highly desirable to enforce long continuous trajectories which automatically bridge occlusions after a re-identification step. The solution proposed in this paper is to connect features in the current image to trajectories which discontinued earlier during the tracking. This is done using a correspondence analysis which is designed for wide baselines and an outlier elimination strategy using the epipolar geometry. The reference to the 3D object points can be used as a new constraint in the bundle adjustment. The feature localization is done using the SIFT detector extended by a Gaussian approximation of the gradient image signal. This technique provides the robustness of SIFT coupled with increased localization accuracy.

Our results show that the reconstruction can be drastically improved and the drift is reduced, especially in sequences with occlusions resulting from foreground objects. In scenarios with large occlusions, the new approach leads to reliable and accurate results while a standard reference method fails.

1 Introduction

Camera motion estimation and simultaneous reconstruction of rigid scene geometry using image sequences is a key technique in many computer vision applications [1,2,3,4,5,6]. The basis for the estimation is the usage of corresponding features which arise from a 3D point being mapped to different camera image planes as shown in Fig. 1. By using a statistical error model which describes the errors in the position of the detected feature points, a Maximum Likelihood estimator can be formulated that simultaneously estimates the camera parameters and the 3D positions of feature points. This joint optimization is called bundle adjustment [7]. It minimizes the distances between the detected feature points and the reprojected 3D points.

Most sequential approaches for structure and motion recovery determine corresponding features in consecutive frames. These correspondences are subsumed

G. Bebis et al. (Eds.): ISVC 2011, Part I, LNCS 6938, pp. 156–167, 2011.

to a trajectory. For small baselines between the cameras, feature tracking methods like KLT [8] are appropriate to obtain stable correspondences. For larger viewpoint changes, features matching methods [9,10,11] have proved impressive performance in determining stable correspondences. The image signal surrounding the feature position is analyzed and distinctive characteristics of this region are extracted for a comparison. These characteristics are assembled to a vector which provides a distinctive representation of the feature. This vector, called descriptor [12], is used to establish correspondences by calculating a similarity measure (L_2 distance) between the current descriptor and the feature descriptors of a second image.

In standard structure and motion recovery approaches, a feature without a correspondence in the previous frame is regarded as a newly appearing object point. If the image feature has been temporarily occluded as shown in Fig. 1, the new object point and the object point that has been generated before the occlusion adopt different 3D positions. As a consequence, errors accumulate and noticeable drift occurs. This problem arises from foreground occlusion as shown in Fig. 2, moving objects, repeated texture, image noise, motion blur, or because tracked points temporarily leave the camera's field of view. For the performance of the bundle adjustment, it is essential to assign reappearing feature points to the correct trajectories and object points.

In case of a closed sequence, the drift can be reduced by enforcing the constraint between the cameras observing the same scene content (i.e. first and last camera view after a complete circuit) [13,14,15]. In [16], the drift is reduced by estimating the transformation between reconstructed 3D point clouds using RANSAC [17]. In [6], broken trajectories caused by occlusion are merged using a combination of localization and similarity constraints of the reprojected object

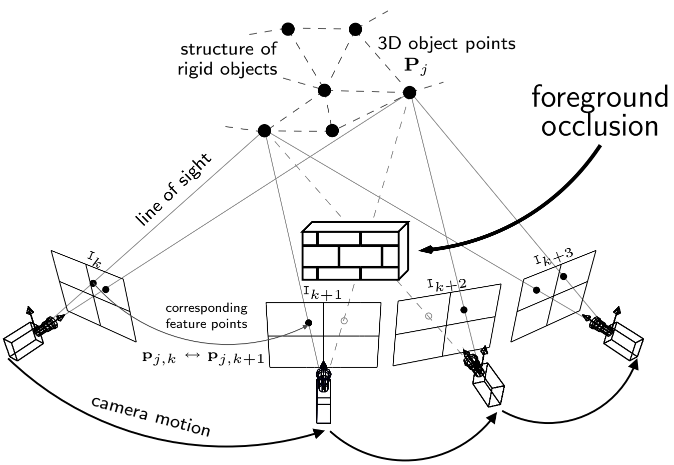
structure of rigid objects
3D object points $\mathbf{P}_j$
foreground occlusion
line of sight
I_k
I_{k+1}
I_{k+2}
I_{k+3}
corresponding feature points
$\mathbf{p}_{j,k} \leftrightarrow \mathbf{p}_{j,k+1}$
camera motion

Fig. 1. The feature trajectories $\mathbf{p}_{j,k}, \mathbf{p}_{j,k+1}, \ldots, \mathbf{p}_{j,k+l}$ are used for structure and motion estimation. Due to foreground occlusion, they discontinue and the corresponding scene content reappears in the images I_{k+2} and I_{k+3}, respectively. A real world example is shown in Fig. 2.

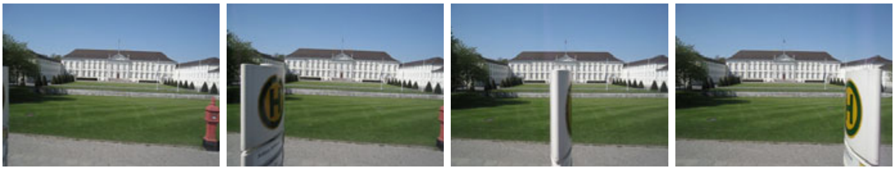

Fig. 2. *Bellevue* sequence: example frames $96, 101, 106, 111$ with temporarily occluded scene content resulting from a foreground object. In standard structure from motion approaches, nearly all feature trajectories discontinue. New object points are generated for scene content after being occluded temporarily.

points in the images. These approaches are only applicable in a post processing step, i.e. after the last frame of the sequence is processed. Furthermore, the estimation has to be accurate before applying the merging. Otherwise it would be impossible to match the cameras, point clouds, or the reprojections of object points. A recent approach [18] directly uses the SIFT descriptor for establishing correspondences in sequential structure and motion recovery. An additional homography constraint for planar features is included to stabilize the tracking using a two-pass matching.

Each of the above mentioned publications neglect to provide an appropriate accuracy evaluation of the reconstructed results. The commonly used reprojection error is not appropriate for the comparison of results with different numbers of constraints in the bundle adjustment. A more accurate solution with more constraints may have a higher reprojection error [16], because it is more likely to find a reconstruction solution for a less constrained system of equations. While the accuracy of the reconstruction increases by enforcing more correct scene relations in the bundle adjustment, the reprojection error increases because of the additional constraints. This important aspect of evaluation is accentuated in this paper.

In our work, the broken trajectories are continued during the tracking by referring a reappearing feature to its last valid occurrence in a previous image. Instead of post-processed merging of different sets of reconstructed object points, our approach immediately continues the trajectory after the occlusion or disturbance of the track. In cases without frame to frame correspondences, the camera recovery can continue without a new initialization. Our feature track retrieval is also beneficial in tracking situations with noise, small occlusions, or repeated texture. The approach is applicable for a live broadcast scenario and may be used to extend any sequential structure and motion recovery method.

The limited localization accuracy of SIFT is addressed by incorporating the feature localization technique presented in [19]. This approach improves the feature localization procedure of SIFT by assuming a Gaussian shape of the feature neighborhood. Hence, the combined approach provides the robustness of SIFT coupled with increased localization accuracy.

This paper provides the following **contributions**:

- a feature tracker which allows to bridge occlusions and therefore generates long feature trajectories yielding to an improved 3D reconstruction
- a detailed analysis of the achieved accuracy of the new tracker using a highly accurate feature localization procedure
- several test scenes demonstrate the superior performance of the proposed combined method

In the following Section 2, the camera motion estimation is briefly explained. The feature localization technique is shown in Section 3. In Section 4, the approach of feature trajectory retrieval is presented. Section 5 shows experimental results using natural image sequences. In Section 6, the paper is concluded.

2 Structure and Motion Recovery

The goal of structure and motion recovery is the accurate estimation of the camera parameters and, simultaneously, of 3D object points of the observed scene [4]. The camera parameters of one camera are represented by the projection matrix $\mathtt{A}_k$ for each image $\mathtt{I}_k$, $k \in [1 : K]$ for a sequence of K image frames. The input data for standard structure and motion estimation consists of corresponding feature points $\mathbf{p}_{j,k}, \mathbf{p}_{j,k+1}$ in consecutive image frames. The accumulation of correspondences over several images is a trajectory $\mathbf{t}_j := (\mathbf{p}_{j,k}, \mathbf{p}_{j,k+1}, \ldots, \mathbf{p}_{j,k+l})$, $j \in [1 : J]$. For each trajectory $\mathbf{t}_j$, a 3D object point $\mathbf{P}_j$ is reconstructed. The 3D-2D correspondence of object and feature point is related by:

$$\mathbf{p}_{j,k} \sim \mathtt{A}_k \mathbf{P}_j \tag{1}$$

where $\sim$ indicates that this is an equality up to scale. The reconstruction starts with an initialization from automatically selected keyframes [20,21]. After computing initial values for the current camera $\mathtt{A}_K$ for frame K, the result is optimized by minimizing the bundle adjustment equation:

$$\epsilon = \sum_{j=1}^{J} \sum_{k=1}^{K} d(\mathbf{p}_{j,k}, \mathtt{A}_k \mathbf{P}_j)^2 \tag{2}$$

The value $r_\epsilon = \sqrt{\frac{\epsilon}{2JK}}$, which is often used for evaluation [14,15,19] is the reprojection error. Object points with small trajectories (length < 3 images) or large reprojection errors are discarded to increase the tracking stability.

For the case of video with small displacements between two frames, feature tracking methods like KLT tend to produce less outliers than feature matching methods. Nevertheless, in this work feature correspondences are established using the SIFT descriptor [9]. It provides more flexibility for reconstructing from images with wide baseline cameras. This also leads to a better interpretability of the accuracy validation of the presented methods.

In the standard structure and motion recovery, features that have no correspondence to the previous frame are considered as new object points. This can lead to significant drift even in short sequences when large occlusions occur.

In this work, this problem is solved using the trajectory retrieval as explained in Section 4. This approach also leads to an increase in the length of the trajectories in general. The second problem is the limited localization accuracy of the SIFT detector, especially for coarse scales [12]. Hence, the localization accuracy is increased using a better approximation of the gradient signal as explained in the following Section 3.

3 Increased Localization Accuracy for SIFT

In [22], it is shown that the feature localization of the SIFT detector can be improved by modifying the gradient signal approximation procedure. The assumption that the image signal around a feature has Gaussian shape is incorporated. Approximately, this assumption can be transferred to the Difference of Gaussians pyramid which is used for the feature localization. Instead of interpolating with a 3D quadric by the original SIFT detector [12], a regression with a Gaussian function is used. The approximation of the selected scale of the Difference of Gaussian pyramid is determined by:

$$G_{\mathbf{p}}(\mathbf{x}) = \frac{v}{\sqrt{|\Sigma|}} \cdot e^{-\frac{1}{2}((\mathbf{x}-\mathbf{x}_0)^{\top}\Sigma^{-1}(\mathbf{x}-\mathbf{x}_0))} \tag{3}$$

using the covariance matrix $\Sigma = \begin{pmatrix} a^2 & b \\ b & c^2 \end{pmatrix}$ and a peak value parameter v. The feature coordinate $\mathbf{x}_0 = (x_0, y_0)$ provides increased localization accuracy. The optimal parameter vector $\mathbf{p} = (x_0, y_0, a, b, c, v)$ is computed using a regression analysis with Levenberg-Marquardt optimization. For details, the reader may refer to [19,22].

4 Feature Trajectory Retrieval

During frame to frame tracking, several trajectories discontinue due to occlusion, repeated texture, or noise in the image signal. By extending the feature comparison to more images, these discontinued trajectories are retrieved if they reappear. These additional constraints are used in the bundle adjustment. To guarantee robust feature matching, the SIFT descriptor [9] is used for the correspondence analysis. In the following, the proposed feature trajectory retrieval is explained using an appropriate memory (Section 4.1) and an outlier elimination method (Section 4.2).

4.1 Trajectory Memory

By using a trajectory memory, the common frame to frame comparison is extended with the correspondences between the current image frame K and frame

$K-L$, $L=2,\ldots,L_{max}$. The memory stores each trajectory $\mathbf{t}_j$ that discontinues and attaches its descriptor for a later retrieval. To guarantee that the stored trajectories correspond to stable object points, only trajectories of a length larger than L_{min} images are considered. For the experiments in this paper, this stability constraint is set to $L_{min}=4$.

Newly appearing features in the current image $\mathtt{I}_K$ with no valid feature correspondence to the previous frame are compared to each trajectory $\mathbf{t}_j$ in the memory using the attached SIFT descriptors. For the matching, the second nearest neighbor search is used [9]. The best match is a candidate to continue the previously discontinued trajectory $\mathbf{t}_j$ as shown in Fig. 3. The candidate has to be verified by using an outlier elimination scheme, which is explained in Section 4.2.

To reduce the computational time, the matching between the current feature and the trajectories is limited to L_{max} past images. The parameter L_{max} controls the size of the trajectory memory. A trajectory in the memory that is older than L_{max} images is deleted. For the experiments in this paper, the memory size is set to $L_{max}=50$ images. To reject false matches, an outlier detection method is applied as explained in the next section.

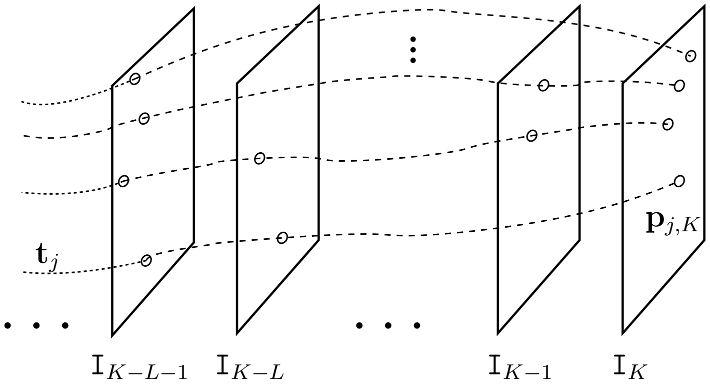

Fig. 3. Schematic feature tracking situation for the proposed Feature Trajectory Retrieval. The newly detected feature $\mathbf{p}_{j,K}$ has a valid corresponding trajectory in image $\mathtt{I}_{K-L}$. Our approach retrieves the corresponding feature trajectory $\mathbf{t}_j$.

4.2 Outlier Elimination

After determining candidates for the resumption of previously discontinued trajectories they have to be verified using the epipolar constraint. Outliers are detected and removed from the correspondence set. If feature correspondences are found between image $\mathtt{I}_K$ and image $\mathtt{I}_{K-L}$, they are evaluated using the RANSAC algorithm [17] together with the epipolar constraint using all valid feature correspondences between $\mathtt{I}_K$ and $\mathtt{I}_{K-L}$. The epipolar constraint is defined by the fundamental matrix $\mathtt{F}$ [23]:

$$\mathbf{p}_{j,k+1}^{\top}\,\mathtt{F}\,\mathbf{p}_{j,k}=0 \quad \forall i \qquad \text{and} \qquad \det(\mathtt{F})=0 \tag{4}$$

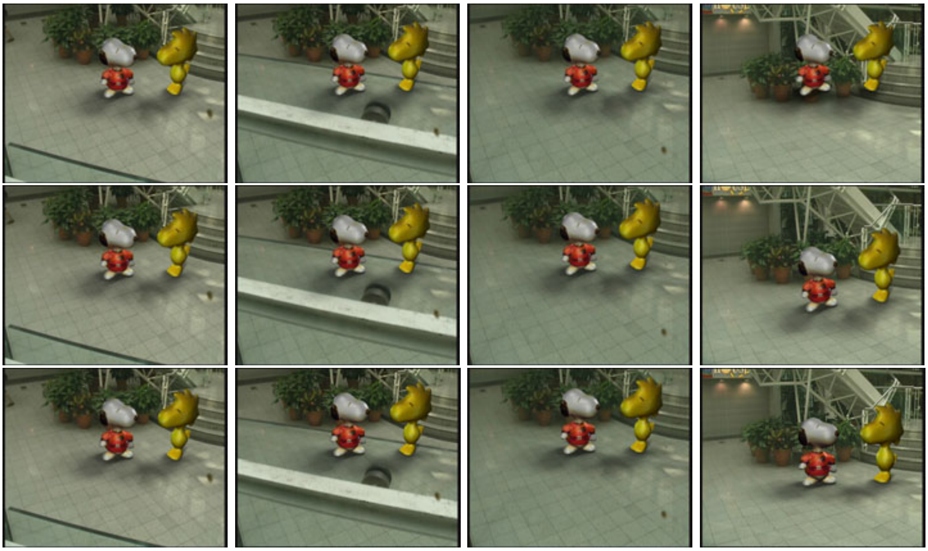

Fig. 4. The *Lift* sequence (frames $19, 26, 33, 83$) with temporarily occluded scene content resulting from a foreground object. In standard structure from motion approaches, all feature trajectories discontinue in frame 30. Thus, the tracking fails (top row). With the presented extensions, the tracking leads to accurate results as shown in the bottom row. Here, augmented objects are integrated in the sequence to demonstrate the accuracy of the results. The center row (SIFT with FTR) shows slight drift of the objects while the top row demonstrates a failure with the SIFT reference method.

To detect the outliers, the RANSAC approach is used evaluating the epipolar distance. The inliers can be referred to previously discontinued trajectories. The outliers remain in the memory for a possibly successful match in the following frames. Using the successfully matched correspondences increases the performance of the bundle adjustment for the sequential camera motion estimation, especially in cases when the number of frame to frame correspondences is low. The bundle adjustment equation (2) is extended with corresponding features before their occlusion. These constraints can be used immediately for the joint optimization in frame K by referring to the already reconstructed and stable object point $\mathbf{P}_j$.

5 Experimental Results

The approaches presented in Section 3 (*Gauss SIFT*) and Section 4 (*Feature Trajectory Retrieval* - FTR) are validated using natural image sequences. Example images are shown in Fig. 2 for the *Bellevue* and in Fig. 4 for the *Lift* sequence[1]. In both cases, the standard structure and motion recovery is strongly influenced by occlusions resulting from foreground objects. The presented methods lead to

[1] Kindly provided by Vicon: http://www.vicon.com/

reliable and accurate results. The FTR provides many useful correspondences to previously discontinued trajectories. Thus, the generation of redundant and error-prone object points is avoided.

For qualitative evaluation, the total number of object points, the object points visible in each frame and the reprojection error r_ϵ are shown in Fig. 5 for the *Bellevue* sequence and in Fig. 6 for the *Lift* sequence. The points in time in which large occlusions occur are marked with t_i. The total number of object points for the FTR is smaller than for the reference method as shown in the top diagram of Fig. 5. This is due to the retrieval of trajectories with corresponding 3D object points. Nevertheless, the number of objects points visible in each frame is higher for the FTR. Thus, more constraints following from larger trajectory lengths are used in the bundle adjustment. This leads to a more accurate reconstruction, although the reprojection error increases for FTR as shown in the bottom

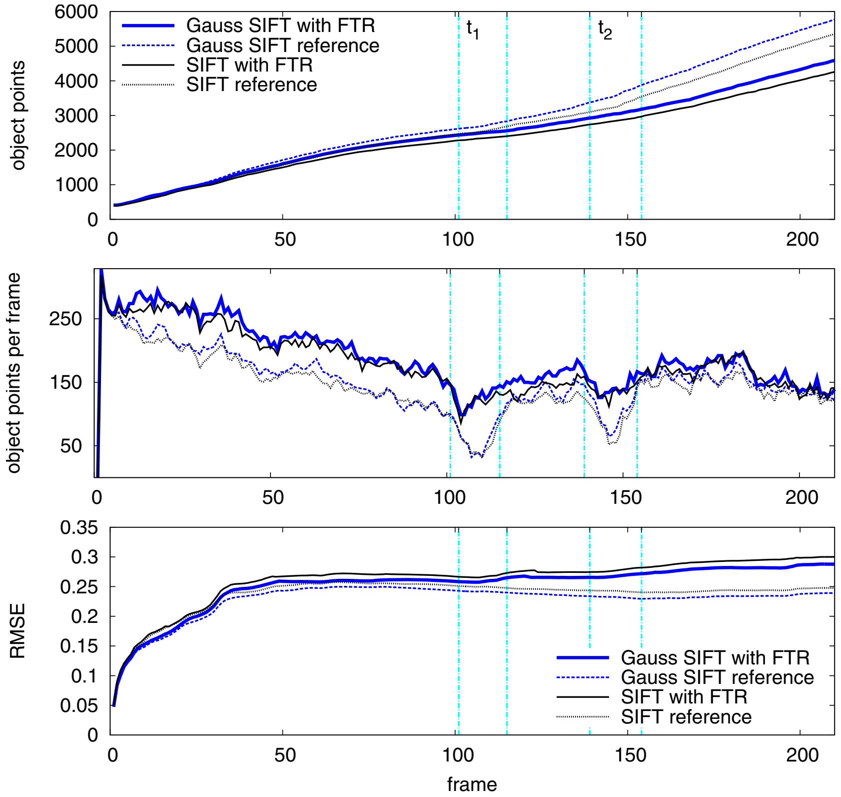

Fig. 5. *Bellevue* reconstruction results, from top to bottom: total number of object points, object points visible in each frame, and reprojection error r_ϵ (*Root Mean Square Error*, RMSE). Due to the Track Retrieval, our method (FTR) provides less object points in total, but more usable constraints in the bundle adjustment. Following from the additional constraints, r_ϵ increases for FTR. Compared to SIFT, the new localization technique (Gauss SIFT) decreases r_ϵ.

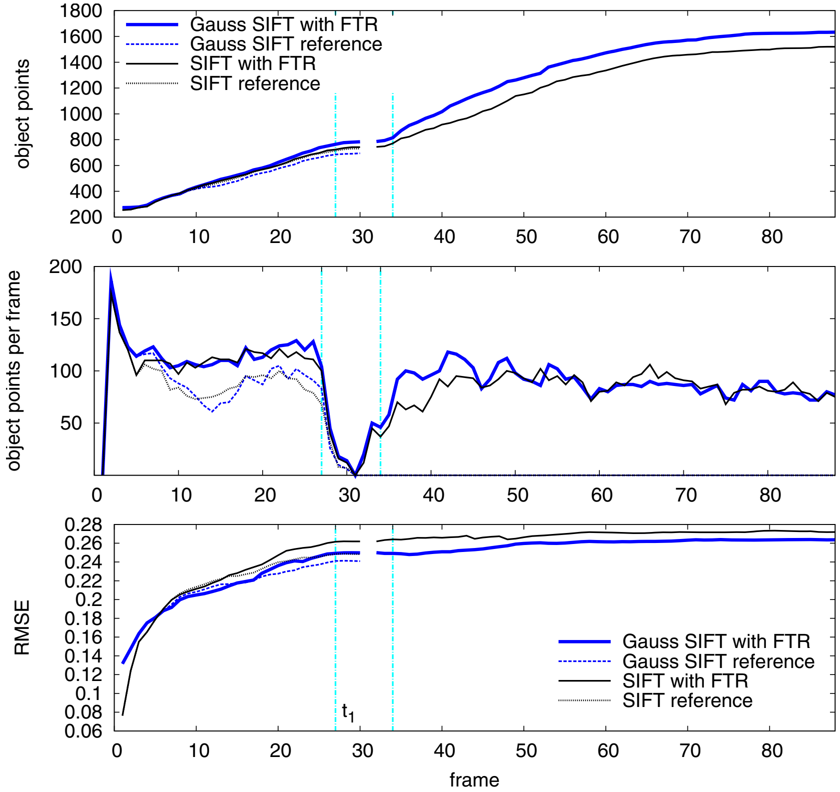

Fig. 6. *Lift* reconstruction results, from top to bottom: total number of object points, object points visible in each frame, and reprojection error r_ϵ (*Root Mean Square Error*, RMSE). Again, our method (FTR) provides more usable constraints in the bundle adjustment. Until frame 30, the more constraints lead to an increase in r_ϵ for FTR. The reference method fails in frame 30. Compared to SIFT, the new localization technique (Gauss SIFT) has lower r_ϵ.

diagrams of Fig. 5 and Fig. 6, respectively. That the scene reconstruction using FTR provides better results is validated by integrating virtual objects to the scene[2] as shown in Fig. 4 for some example frames.

Due to an insufficient number of valid correspondences, the reference method fails for the *Lift* sequence in frame 30. Interestingly, the FTR leads to more object points per frame for the sequence parts without occlusion, too. We can infer that the method also increases the performance in occurrence of noise and repeated texture, which is present in many scene parts in the *Bellevue* sequence (windows, grass texture).

[2] The video can be downloaded at:
`http://www.tnt.uni-hannover.de/staff/cordes/`

The feature localization technique using the Gaussian regression function (*Gauss SIFT with FTR*) leads to better results in any case. Compared to the localization of SIFT (*SIFT with FTR*), it results in more object points, more object points per frame and, even for the higher amount of object points, a smaller reprojection error. Usually, the reprojection error increases when more constraints are added in the bundle adjustment, because it is more likely to find a reconstruction solution for a less constraint system of equations. Here, the results provide decreased reprojection error *and* more object points.

The approaches are tested using the application scenario of integrating virtual objects into the image sequence. To validate the reconstruction accuracy, for the *Lift* sequence two static objects are integrated, which is shown in Fig. 4. Due to an accurately estimated camera path, the integration is convincing throughout the sequence using the combination of FTR and Gauss SIFT. The objects show a drift using the FTR and the standard SIFT feature localization technique. The structure and motion estimation fails in frame 30 for the reference method.

6 Conclusion

We present an improved sequential structure and motion recovery approach. Discontinued feature trajectories are retrieved using the distinctive SIFT descriptor and a correspondence analysis for nonconsecutive image frames. The retrieved correspondences between features in the current image and previously discontinued trajectories can be immediately used in the bundle adjustment of the current frame. The method leads to longer trajectories and avoids generating redundant and error-prone 3D object points. As a result, the bundle adjustment performance is increased.

To compensate for the limited localization accuracy of the SIFT detector, an improved localization method is applied. It uses a Gaussian approximation of the image signal gradient instead of the interpolation with a 3D quadric used by the original SIFT.

A problem using the reprojection error for comparing reconstructions from different numbers of scene constraints is revealed: while the reconstruction improves using more scene relations, the reprojection error increases because the additional constraints limit the possible solutions of the system of equations used for the bundle adjustment.

The performance is validated using natural image sequences with temporal occlusions resulting from foreground objects. While the results of the reference method suffer from broken trajectories, the presented approach using the combination of highly-accurate feature localization and feature trajectory retrieval shows no drift. This is demonstrated by integrating augmented objects to the video. The presented extensions are applicable to any sequential structure and motion recovery algorithm.

References

1. Frahm, J.M., Pollefeys, M., Lazebnik, S., Gallup, D., Clipp, B., Raguram, R., Wu, C., Zach, C., Johnson, T.: Fast robust large-scale mapping from video and internet photo collections. Journal of Photogrammetry and Remote Sensing (ISPRS) 65, 538–549 (2010)
2. Hasler, N., Rosenhahn, B., Thormählen, T., Wand, M., Seidel, H.P.: Markerless motion capture with unsynchronized moving cameras. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2009)
3. van den Hengel, A., Dick, A., Thormählen, T., Ward, B., Torr, P.H.S.: Videotrace: rapid interactive scene modelling from video. In: ACM SIGGRAPH 2007 papers. SIGGRAPH 2007, vol. (86). ACM, New York (2007)
4. Pollefeys, M., Gool, L.V.V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. International Journal of Computer Vision (IJCV) 59, 207–232 (2004)
5. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the world from internet photo collections. International Journal of Computer Vision (IJCV) 80, 189–210 (2008)
6. Thormählen, T., Hasler, N., Wand, M., Seidel, H.P.: Registration of sub-sequence and multi-camera reconstructions for camera motion estimation. Journal of Virtual Reality and Broadcasting 7 (2010)
7. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment - a modern synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) ICCV-WS 1999. LNCS, vol. 1883, pp. 298–372. Springer, Heidelberg (2000)
8. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: International Joint Conference on Artificial Intelligence (IJCAI), pp. 674–679 (1981)
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (IJCV) 60, 91–110 (2004)
10. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: British Machine Vision Conference (BMVC), vol. 1, pp. 384–393 (2002)
11. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 27, 1615–1630 (2005)
12. Brown, M., Lowe, D.G.: Invariant features from interest point groups. In: British Machine Vision Conference (BMVC), pp. 656–665 (2002)
13. Engels, C., Fraundorfer, F., Nistér, D.: Integration of tracked and recognized features for locally and globally robust structure from motion. In: VISAPP (Workshop on Robot Perception), pp. 13–22 (2008)
14. Fitzgibbon, A.W., Zisserman, A.: Automatic camera recovery for closed or open image sequences. In: Burkhardt, H.-J., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, pp. 311–326. Springer, Heidelberg (1998)
15. Liu, J., Hubbold, R.: Automatic camera calibration and scene reconstruction with scale-invariant features. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Nefian, A.V., Gopi, M., Pascucci, V., Zara, J., Molineros, J., Theisel, H., Malzbender, T. (eds.) ISVC 2006. LNCS, vol. 4291, pp. 558–568. Springer, Heidelberg (2006)
16. Cornelis, K., Verbiest, F., Van Gool, L.: Drift detection and removal for sequential structure from motion algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 26, 1249–1259 (2004)

17. Fischler, R.M.A., Bolles, C.: Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. Communications of the ACM 24, 381–395 (1981)
18. Zhang, G., Dong, Z., Jia, J., Wong, T.T., Bao, H.: Efficient non-consecutive feature tracking for structure-from-motion. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part V. LNCS, vol. 6315, pp. 422–435. Springer, Heidelberg (2010)
19. Cordes, K., Müller, O., Rosenhahn, B., Ostermann, J.: Bivariate feature localization for sift assuming a gaussian feature shape. In: Bebis, G., Boyle, R.D., Parvin, B., Koracin, D., Chung, R., Hammoud, R.I., Hussain, M., Tan, K.H., Crawfis, R., Thalmann, D., Kao, D., Avila, L. (eds.) ISVC 2010. LNCS, vol. 6453, pp. 264–275. Springer, Heidelberg (2010)
20. Thormählen, T., Broszio, H., Weissenfeld, A.: Keyframe selection for camera motion and structure estimation from multiple views. In: Pajdla, T., Matas, J. (eds.) ECCV 2004, Part I. LNCS, vol. 3021, pp. 523–535. Springer, Heidelberg (2004)
21. Torr, P.H.S., Fitzgibbon, A.W., Zisserman, A.: The problem of degeneracy in structure and motion recovery from uncalibrated image sequences. International Journal of Computer Vision (IJCV) 32, 27–44 (1999)
22. Cordes, K., Müller, O., Rosenhahn, B., Ostermann, J.: Half-sift: High-accurate localized features for sift. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Workshop on Feature Detectors and Descriptors: The State Of The Art and Beyond, pp. 31–38 (2009)
23. Hartley, R.I., Zisserman, A.: Multiple View Geometry, 2nd edn. Cambridge University Press, Cambridge (2003)